Precedence based speech segregation in bilateral cochlear implant users

Shaikat Hossain,[1] Vahid Montazeri,[1] Peter F. Assmann,[1] and Ruth Y. Litovsky[2]

[1]*School of Behavioral and Brain Sciences, University of Texas at Dallas, Richardson Texas 75083-0688, USA*

[2]*Department of Communication Disorders and Waisman Center, University of Wisconsin, Madison Wisconsin 53706, USA*

shaikat@utdallas.edu, vahid.montazeri@utdallas.edu, assmann@utdallas.edu, litovsky@waisman.wisc.edu

Abstract: The precedence effect (PE) enables the perceptual dominance by a source (lead) over an echo (lag) in reverberant environments. In addition to facilitating sound localization, the PE can play an important role in spatial unmasking of speech. Listeners attending to binaural vocoder simulations with identical channel center frequencies and phase demonstrated PE-based benefits in a closed-set speech segregation task. When presented with the same stimuli, bilateral cochlear implant users did not derive such benefits. These findings suggest that envelope extraction in itself may not lead to a breakdown of the PE benefits, and that other factors may play a role.

[Q-JF]

Date Received: September 9, 2015 **Date Accepted:** December 2, 2015

1. Introduction

Cochlear implant (CI) devices have been successful in providing speech reception abilities to many severe-to-profoundly deafened patients in quiet settings. However, CI users continue to face great difficulties in acoustically adverse listening conditions, such as classrooms and work environments. Under such circumstances, normal hearing (NH) listeners often receive significant benefits from binaural cues, because interaural time and level differences in sounds arriving at the two ears can facilitate a release from masking. Studies have shown that bilateral CI (BiCI) users perform better than unilateral CI users on speech recognition in noise and sound localization tasks (Litovsky *et al.*, 2009). However, BiCI users continue to experience difficulty understanding speech and localizing sounds in noisy and reverberant environments (Kerber and Seeber, 2013). The limited benefit received from BiCIs may be due to the lack of synchronization between the two processors and the spread of electrical current within each cochlea, factors that degrade interaural cues and subsequently reduce binaural sensitivity (Kan and Litovsky, 2015). While interaural level differences and interaural timing differences in the envelopes can be provided by current devices, the primary benefit derived from having two implants is largely based on contrasting monaural cues from each ear (i.e., the "better ear effect"). However, additional benefit could potentially be derived from restoring binaural unmasking mechanisms found in NH listeners, which might lead to substantial improvements in speech understanding in noise and in reverberant environments.

This study focuses on the precedence effect (PE), a particularly important auditory mechanism in which the auditory system assigns preferential weighting to the directional cues carried by the first arriving sound and minimizes the weighting assigned to later arriving reflections/echoes. The PE is thus thought to facilitate directional hearing in acoustically reverberant environments (Litovsky *et al.*, 1999; Brown *et al.*, 2015b). Fusion of simulated source (lead) and echo (lag) sounds, and localization dominance of the leading sound are two aspects of the PE. In addition to aiding sound localization, the PE has been shown to contribute to the spatial unmasking of speech for NH listeners. For conditions in which a masker is co-located with the target, the PE can induce perceived spatial separation, which is effective at producing release from masking (Freyman *et al.*, 1999).

Although spatial hearing has been investigated in BiCI users, little is known about the PE in this population. Studies have demonstrated reduced fusion of lead and lag in vocoder-based simulations (Seeber and Hafter, 2011). Recent findings have also shown that BiCI users are capable of experiencing the PE when using direct stimulation on a single electrode pair (Brown *et al.*, 2015a; Agrawal *et al.*, 2008). However,

Agrawal *et al.* (2008) reported that the PE was weak or absent for the same BiCI users listening in sound field through their clinical processors. Furthermore, the relationship between the PE and the speech segregation abilities of BiCI users remains unclear. The present study takes preliminary steps towards evaluating CI recipients' listening performance in a PE-based speech segregation task established by Freyman *et al.* (1999). Based on previous research demonstrating the importance of interaural temporal synchronization and optimal pitch-matching for the effective encoding of binaural cues (Long *et al.*, 2003; Poon *et al.*, 2009; Kan *et al.*, 2013), we expect that BiCI users tested using their clinical processors will exhibit a weak PE benefit. Furthermore, we hypothesize that NH listeners in a vocoder simulation condition with temporally aligned channels with matched center frequencies will more effectively encode interaural cues and subsequently exhibit greater PE-based speech segregation benefits. The findings from the present study seek to improve our understanding of basic auditory function as well as inform design of future signal processing strategies aimed at coordinating CIs in BiCI users.

2. Methods

2.1 Participants

Six post-lingually deafened BiCI users were recruited for participation (one male, five females). All wore Cochlear Corporation Nucleus devices and ranged in age between 55 and 80 years (mean age of 64 years). In addition to the BiCI listeners, ten NH native English speakers (three males, seven females), who ranged in age between 18 and 28 years (mean age of 22 years), were recruited for participation in this study. These participants were students at the University of Texas at Dallas who were compensated with course credit for their time. All ten listeners had their pure tone thresholds tested at 20 dB hearing level (HL) at octave frequencies from 250 to 8000 Hz in both ears. All procedures involving human subjects were reviewed and approved by the University of Texas Institutional Review Board.

2.2 Stimuli

The stimuli used in the study were Coordinate Response Measure (CRM) sentences spoken by four male talkers. Sentences were of the form "Ready (name) go to (color) (number) now." Target sentences always contained the name "Baron," with the masker sentence containing a different name. Talker, color, and number varied across masker and target sentences with the restriction that any particular combination of words were not repeated across any of the target or masker sentences. Testing was conducted with two types of maskers: speech and speech-shaped noise (SSN). The spectrum of the SSN stimulus was calculated by taking the average of the log-magnitude spectra of all the phrases in the CRM corpus and used to design a finite impulse response filter which was used to impose the averaged spectral envelope of the speech corpus onto Gaussian white noise. Sentences were root-mean-square equalized and scaled to create three different signal-to-noise ratio (SNR) conditions (−4, 0, +4 dB) in reference to the target speech.

2.3 Procedure

Participants were seated in front of a computer monitor in a soundproofed booth and were asked to complete a closed-set word recognition task while listening through Bose 201 loudspeakers (for the BiCI users) or through Audio-Technica ATH-M45 headphones (for the NH listeners) in a soundproof booth. A graphical user interface generated in MATLAB allowed participants to select their responses indicating the color and number in the target sentence by clicking on the corresponding buttons with a mouse. The range of SNRs included in the experiment was carefully chosen through pilot testing using vocoder simulations to find the optimal range of performance on the task. In the free field condition, BiCI users were tested in free field in a soundproof booth with the subjects seated one meter from both loudspeakers. Loudspeakers were placed at 0° (front) and 60° (right). Target and masker stimuli were presented from one of the following configurations. (A) Both sounds from the front loudspeaker (F_F). (B) target from the front and masker from the right (F_R). (C) Target from the front and masker from the right, plus a delayed copy of the masker added back to the front loudspeaker after a 4 ms delay (F_RF). The 4 ms delay falls within the echo threshold of running speech (Litovsky *et al.*, 1999), and this latter configuration has been found to produce spatial unmasking due to the perceived spatial separate between the target and masker (Freyman *et al.*, 1999). That is, due to the PE, the masker is perceived to be from the right, even though one copy of the masker is co-located with the target. The protocol

was based on the methodology of Freyman *et al.* (1999). Word recognition percent correct scores were calculated by taking the number of words identified correctly divided by the total number of words in a given sentence, averaged across all tokens for a particular condition.

2.4 Processing

For the NH listeners, additional processing was applied. Non-individualized head-related transfer functions (HRTFs) were used to create a virtual auditory space for headphone presentation. The stimuli were first processed with the same set of HRTFs were used in the Brungart *et al.* (2005) study, simulating source locations at 0° and 60° in the azimuthal plane for the three spatial configurations. For the F_F configuration, both target and masker were processed with HRTFs measured at 0° azimuth. For the F_R configuration, the target was processed with left and right HRTFs measured at 0° and the masker was processed with the HRTFs measured at 60°. For the F_RF configuration, target and masker were processed as in the F_R configuration, with an additional copy of the masker delayed by 4 ms, processed by HRTFs at 0° added onto the stimulus.

The stimuli were subsequently processed by a binaural eight-channel sine-excited vocoder based on the implementation used by Fu *et al.* (2004). First, the signal was processed through a high-pass pre-emphasis filter with a cutoff of 1200 Hz and a slope of −6 dB/octave. The input frequency range of 200–7000 Hz was then divided into eight frequency bands, using fourth order Butterworth filters. The distribution of these filters was based on the Greenwood function, using the same corner frequencies as in Fu *et al.* (2004). The temporal envelope of each band was extracted using half-wave rectification and low-pass filtering with a cutoff frequency of 160 Hz. The extracted envelopes were then used to modulate the amplitude of the corresponding sine-wave. Sine carriers across both ears were identical in frequency and were phase-aligned.

3. Results

Given that the stimuli used in testing the BiCI users and NH listeners differed due to mode of presentation (loudspeakers versus HRTF-processed stimuli over headphones), two separate analyses were conducted. A three-factor repeated-measures analysis of variance (ANOVA) was performed on percent correct word recognition scores collected from the BiCI users. The factors were masker type (two levels: speech and SSN), configuration (three levels: F_F, F_R, and F_RF), and SNR (three levels: −4, 0, and 4 dB). There were main effects of masker type [$F(1,5) = 83.60$, $p < 0.01$], configuration [$F(2,10) = 22.91$, $p < 0.01$], and SNR [$F(2,10) = 207.06$, $p < 0.01$]. All two-way interactions were significant ($p < 0.05$) and there was a significant three-way interaction [$F(4,20) = 3.78$, $p < 0.05$]. Figure 1 shows performance of BiCI users. A three-factor repeated measures ANOVA was also performed on percent correct word recognition scores collected from the NH users with masker type (two levels: speech and SSN), configuration (three levels: F_F, F_R, and F_RF), and SNR (three levels: −4, 0, and 4 dB) as factors. There were main effects of masker type [$F(1,9) = 20.02$, $p < 0.05$], configuration [$F(2,18) = 199.11$, $p < 0.01$], and SNR [$F(2,18) = 545.81$, $p < 0.01$]. All two-way interactions were significant ($p < 0.01$) and there was a significant three-way

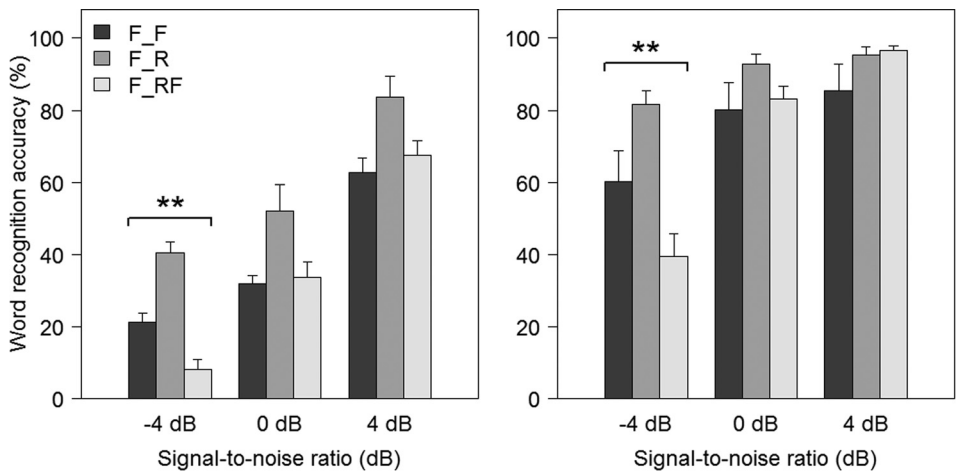

Fig. 1. Performance of BiCI users with speech maskers (left) and SSN (right).

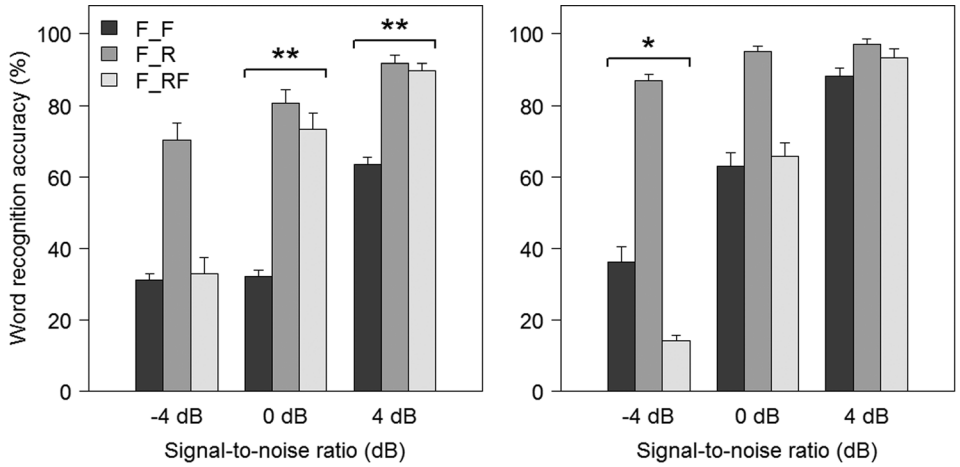

Fig. 2. Performance of NH listeners with speech maskers (left) and SSN (right).

interaction [$F(4,36) = 11.76$, $p < 0.01$]. Figure 2 shows performance of the NH listeners. Pairwise t-tests were performed to determine whether there were differences between the F_F and F_RF configurations for both the NH and CI data sets. Tests were conducted across all levels of masker type crossed with SNR, leading to a total of six comparisons (2 levels of masker type $\times 3$ levels of SNR) for each dataset (BiCI and vocoder simulation). The Bonferroni method was used to correct the alpha criterion for multiple comparisons. The F_RF configuration failed to lead to any improvements in word recognition scores for the BiCI users for either of the two masker types. At -4 dB SNR, a reduction in scores in the F_RF configuration as compared to the F_F configuration was significant for both speech ($t = 4.97$, $p < 0.01$) and SSN maskers ($t = 3.63$, $p < 0.01$). The NH listeners exhibited higher scores in the F_RF configuration as compared to the F_F configuration at 0 dB ($t = 7.99$, $p < 0.01$) and 4 dB ($t = 7.78$, $p < 0.01$) for the speech maskers. For the SSN condition, the scores in the F_RF configuration were significantly lower than the scores in the F_F configuration at -4 dB SNR ($t = 4.97$, $p < 0.05$).

We next calculated the proportion of intrusion errors (substituting the masker in place of target) as a measure of informational masking (Kidd *et al.*, 1994). There were strong negative linear correlations between difference scores (F_RF – F_F) in word recognition accuracy and difference scores in intrusion error rates for the NH listeners ($r = -0.95$ at -4 dB, $r = -0.94$ at 0 dB, and $r = -0.88$ at $+4$ dB SNR; $N = 10$; $p < 0.01$) and somewhat lower correlations for the BiCI users ($r = -0.83$, $p < 0.05$, at -4 dB, $r = -0.89$, $p < 0.05$, at 0 dB, and $r = -0.65$, $p > 0.1$, at $+4$ dB SNR; $N = 6$). A two-factor ANOVA was performed on the proportion of intrusion errors of the BiCI users with configuration (three levels: F_F, F_R, F_RF) and SNR (-4, 0, 4 dB) as factors. There was a main effect of configuration [$F(2,10) = 5.51$, $p < 0.05$] and SNR [$F(2,10) = 44.39$, $p < 0.01$]. In addition, there was a two-way interaction between configuration and SNR [$F(4,20) = 6.75$, $p < 0.01$]. A two-factor ANOVA was also performed on the intrusion errors of the NH listeners with configuration (three levels: F_F, F_R, F_RF) and SNR (-4, 0, 4 dB) as factors. There was a main effect of configuration [$F(2,18) = 39.11$, $p < 0.01$] and SNR [$F(2,18) = 72.07$, $p < 0.01$]. In addition, there was a two-way interaction between configuration and SNR [$F(4,36) = 12.96$, $p < 0.01$]. Pairwise t-tests were performed to determine whether there were differences in the occurrence of intrusion errors between the F_F and F_RF configurations within each SNR. Bonferroni correction was used to adjust the alpha criterion for multiple comparisons in evaluating the p-values. For the BiCI users, there was not a significant reduction in intrusion errors at any of the SNR values. An increase in intrusion errors in the F_RF configuration compared to the F_F configuration at -4 dB SNR approached significance ($t = 1.94$, $p = 0.055$). For the NH listeners, there was a significant reduction in intrusion errors in the F_RF configuration at 0 dB ($t = 6.58$, $p < 0.01$) and at 4 dB ($t = 4.88$, $p < 0.05$).

4. Discussion

BiCI users tested in this study did not derive benefit from the PE in the speech segregation task, as indicated by the lack of improvement from F_F to F_RF configurations. In fact, they exhibited the poorest performance in the F_RF configuration at the lowest SNR (-4 dB) for both speech and SSN maskers. One interpretation is that, unlike a normal auditory system where a simulated echo can be largely ignored, the same

echo may be detrimental for BiCI users listening through their clinical processors at low SNRs. This result is consistent with previous findings of CI user's reduced ability to recognize speech in reverberant environments using current sound processing strategies (Kerber and Seeber, 2013). Although the simulated echo used here is not the same as the reverberation that occurs in everyday listening situations, the lagging sound occurs at a short delay, is not perceived as a separate sound by the listeners, yet is detrimental to speech understanding.

One possible explanation for the reduced ability of BiCI users to process early reflections is that clinical processors may introduce distortions to interaural cues which subsequently reduce the effectiveness of the PE at minimizing the influence of echoes. These distortions could result from a number of factors such as (1) interaural temporal jitter resulting from the two CI processors running on independent clocks, (2) lack of matched stimulation of electrodes in the two cochlea at anatomical locations that activate similar frequency regions, and (3) variation in microphone characteristics. For example, mismatch between the ears in the place of stimulation mismatch has been shown to produce poorer fusion and lateralization, which might explain the decrease in binaural sensitivity in BiCI users (Kan *et al.*, 2015; Kan *et al.*, 2013). It would follow that the vocoder simulation condition, which implicitly temporally aligned channels with matched center frequencies across the ears, may have led to the veridical encoding of interaural cues and, thus, an improved capacity to experience the PE (Kan and Litovsky, 2015).

Unlike the BiCI users, data from the vocoder simulation with NH listeners showed higher word recognition scores in the F_RF as compared to the F_F configuration at the more favorable SNRs (0 and 4 dB). These findings support the interpretation that, despite lack of temporal fine structure in the signal, the PE was effective at improving the ability of listeners to segregate sounds. This benefit was exhibited for the speech maskers but not for the SSN, consistent with data previously collected from NH listeners (Brungart *et al.*, 2005). The specificity of the benefit may relate to temporal modulations in speech maskers not present in the SSN that provided listeners with more opportunities to "glimpse" the target signals. Future work on a glimpsing-based explanation for this benefit could help establish whether fluctuations in the temporal envelope of the masker are related to the benefit.

Glimpsing the temporal peaks of the target within the valleys of the masker provides an explanation that is based purely on the acoustic characteristics of the sound. Brungart *et al.* (2005) put forward a complementary explanation framed in terms of informational masking. The release from masking resulting from the PE could be related to a reduction in the cognitive load of having to parse and process the semantic content of two concurrent sentences. BiCI users faced greater difficulty with the speech maskers than the SSN, which could be attributed to a perceptual inability to separate target/masker streams. For the NH listeners there were significant reductions in intrusion errors in F_RF as compared to F_F configuration at 0 and 4 dB SNR. These are the conditions where we found significant PE benefits in the word recognition scores (F_RF > F_F). This pattern was not seen in the BiCI data. Furthermore, the increase in intrusion errors in the F_RF configuration at −4 dB SNR suggests that early reflections may be particularly detrimental to BiCI users at low SNRs. The lack of benefit derived from the PE for BiCI users could be related to the persistence of intrusion errors in the F_RF configuration. This idea is supported by the presence of significant correlations between word recognition accuracy improvement and the reduction in intrusion errors in the F_RF configuration.

BiCI users tested in this experiment did not experience PE-related benefits in speech segregation. Factors contributing to the absence of PE-based benefit for the BiCI users may include microphone characteristics or the lack of coordination between processors. Timing delays introduced by individual processors running on different clocks and mismatched frequency-to-place maps between ears may have led to distortions in interaural cues that are critical for the PE. We are currently testing BiCI users using a coordinated MAP to test the hypothesis that temporal synchronization and optimal pitch-matching are required for the effective encoding of interaural cues and facilitation of PE-based speech segregation benefits.

Acknowledgments

This publication is dedicated to the memory of Professor Philipos Loizou. The work was supported by NIH-NIDCD [Grant No. R01 DC010494, P. Loizou principal investigator (PI) until 2011; John Hansen current PI] and Grant No. R01 DC003083 (R. Litovsky PI).

References and links

Agrawal, S. (**2008**). "Spatial hearing abilities in adults with bilateral cochlear implants," Doctoral thesis, University of Wisconsin, Madison.

Brown, A., Jones, H., Kan, A., Thakkar, T., Stecker, G. C., Goupell, M., and Litovsky, R. (**2015a**). "Evidence for a neural source of the precedence effect in sound localization," J. Neurophysiol. **114**(5), 2991–3001.

Brown, A. D., Stecker, G. C., and Tollin, D. J. (**2015b**). "The precedence effect in sound localization," J. Assoc. Res. Otolaryngol. **16**(1), 1–28.

Brungart, D. S., Simpson, B. D., and Freyman, R. L. (**2005**). "Precedence-based speech segregation in a virtual auditory environment," J. Acoust. Soc. Am. **118**, 3241–3251.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.

Fu, Q. J., Chinchilla, S., and Galvin, J. J. (**2004**). "The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users," J. Assoc. Res. Otolaryngol. **5**(3), 253–260.

Kan, A., and Litovsky, R. Y. (**2015**). "Binaural hearing with electrical stimulation," Hear. Res. **322**, 127–137.

Kan, A., Litovsky, R. Y., and Goupell, M. J. (**2015**). "Effects of interaural pitch-matching and auditory image centering on binaural sensitivity in cochlear-implant users," Ear. Hear. **36**(3), 62–68.

Kan, A., Stoelb, C., Litovsky, R. Y., and Goupell, M. J. (**2013**). "Effect of mismatched place-of-stimulation on binaural fusion and lateralization in bilateral cochlear-implant users," J. Acoust. Soc. Am. **134**(4), 2923–2936.

Kerber, S., and Seeber, B. U. (**2013**). "Localization in reverberation with cochlear implants: Predicting performance from basic psychophysical measures," J. Assoc. Res. Otolaryngol. **14**(3), 379–392.

Kidd, G., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, H. S. (**1994**). "Reducing informational masking by sound segregation," J. Acoust. Soc. Am. **95**, 3475–3480.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). "The precedence effect," J. Acoust. Soc. Am. **106**, 1633–1654.

Litovsky, R. Y., Parkinson, A., and Arcaroli, J. (**2009**). "Spatial hearing and speech intelligibility in bilateral cochlear implant users," Ear Hear. **27**(6), 714–731.

Long, C. J., Eddington, D. K., Colburn, H. S., and Rabinowitz, W. M. (**2003**). "Binaural sensitivity as a function of interaural electrode position with a bilateral cochlear implant user," J. Acoust. Soc. Am. **114**, 1565–1574.

Poon, B. B., Eddington, D. K., Noel, V., and Colburn, H. S. (**2009**). "Sensitivity to interaural time difference with bilateral cochlear implants: Development over time and effect of interaural electrode spacing," J. Acoust. Soc. Am. **126**, 806–815.

Seeber, B. U., and Hafter, E. R. (**2011**). "Failure of the precedence effect with a noise-band vocoder," J. Acoust. Soc. Am. **129**, 1509–1521.